DOCUMENT RESUME

ED 251 487 TM 840 763

AUTHOR Bennett, Randy Elliot; And Others
TITLE The Test Performance of Handicapped People. Studies of Admissions Testing and Handicapped People, Report No. 2.
INSTITUTION College Entrance Examination Board, New York, N.Y.; Educational Testing Service, Princeton, N.J.; Tennessee Univ., Knoxville. Dept. of Distributive Education.
REPORT NO ETS-RR-84-32
PUB DATE Aug 84
NOTE 54p.; For Report No. 1, see TM 840 757.
AVAILABLE FROM Educational Testing Service, Research Publications R-116, Princeton, NJ 08541.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *College Entrance Examinations; *Disabilities; Hearing Impairments; Higher Education; Learning Disabilities; Physical Disabilities; *Test Reliability; *Test Results; *Test Validity; Visual Impairments
IDENTIFIERS *Act Assessment; *Scholastic Aptitude Test

ABSTRACT

The purpose of this paper is to document existing research on the test performance of handicapped people with respect to admissions and other similar tests. In addition, the psychometric characteristics of these tests when used with handicapped people are reported. Though much more data needs to be collected, several observations are made. Handicapped students as a general group perform appreciably lower than national norms on the Scholastic Aptitude Test (SAT) and ACT Assessment. Of the four specific disability groups discussed, the admissions test performance of physically handicapped and of visually impaired examinees is most similar to the nondisabled population. The performance of learning disabled individuals ranks third among the specific disability groups, and is appreciably lower than national norms. Hearing impaired students perform the least well of all disabled groups on admissions tests. The limited data on the reliability of cognitive tests for handicapped examinees show no trustworthy differences in measurement precision between disabled and nondisabled populations. The results of the small number of validity investigations that have been conducted also show no dependable differences across popuulations. Data on the ACT Assessment and the SAT generally support the validity of these measures as equivalent predictors of college performance. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED251487

THE TEST PERFORMANCE OF HANDICAPPED PEOPLE

Randy Elliot Bennett
Marjorie Ragosta
and
Lawrence J. Stricker

August 1984

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. Wadsworth Miller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Report No. 2
Studies of Admissions Testing and Handicapped People
A Project Sponsored by

College Entrance Examination Board
Educational Testing Service
Graduate Record Examinations Board

TM 840 763

Studies of Admissions Testing and Handicapped People

Most admissious testing programs have long made accommodations for handicapped examinees, though practices have varied across programs and limited research has been undertaken to evaluate such test modifications. Regulations under Section 504 of the Rehabilitation Act of 1973 impose new requirements on institutional users, and indirectly on admissions test sponsors and developers, in order to protect the rights of handicapped persons. The Regulations have not been strictly enforced since many have argued that they conflict with present technical capabilities of test developers. In 1982, a Panel appointed by the National Research Council released a detailed report and recommendations calling for research on the validity and comparability of scores for handicapped persons.

Due to a shared concern for these issues, College Board, Educational Testing Service, and Graduate Record Examinations Board initiated a series of studies in June 1983. The primary objectives are:

To develop an improved base of information concerning the testing of handicapped populations.

To evaluate and improve wherever possible the accuracy of assessment for handicapped persons, especially test scaling and predictive validity.

To evaluate and enhance wherever possible the fairness and comparability of tests for handicapped and nonhandicapped examinees.

This is one of a series of reports on the project, which will continue through 1986.

ETS Research Report RR-84-32

The Test Performance of Handicapped People

Randy Elliot Bennett

Marjorie Ragosta

and

Lawrence J. Stricker[1]

Educational Testing Service

Princeton, NJ 08541

August 1984

Copyright © 1984 by Educational Testing Service. All rights reserved.

The ETS logo, Educational Testing Service, GRE, and Graduate Record Examinations Board, are registered trademarks of Educational Testing Service.

College Board and the acorn logo are registered trademarks of the College Entrance Examination Board.

Abstract

This paper is the second in a series of reports emanating from a four-year research program meant to further knowledge of college and graduate admissions testing and handicapped people. The purpose of this paper is to document existing research on the test performance of handicapped people with respect to admissions and other similar tests. In addition, the psychometric characteristics of these tests when used with handicapped people are reported.

Though much more data need to be collected, several observations are made. Handicapped students as a general group perform appreciably lower than national norms on the SAT and ACT Assessment (by about .5 standard deviations). Of the four specific disability groups discussed, the admissions test performance of physically handicapped and of visually impaired examinees is most similar to the nondisabled population. The admissions test performance of learning disabled individuals ranks third among the specific disability groups. With limited exceptions, the performance of this group is appreciably lower than national norms (generally by at least .5 standard deviations). Hearing impaired students perform the least well of all disabled groups on admissions tests.

The limited data on the reliability of cognitive tests for handicapped examinees show no trustworthy differences in measurement precision between disabled and nondisabled populations. The results of the small number of validity investigations that have been conducted also show no dependable differences across populations. Data on the ACT Assessment and

the SAT generally support the validity of these measures as equivalent predictors of college performance in selected disabled student groups and the nonhandicapped population.

The lack of data on test performance, reliability, and validity prohibits drawing definitive conclusions at this time. Further research is necessary before the fairness of the admissions testing process for handicapped examinees can be soundly evaluated.

Acknowledgements

Appreciation is extended to those individuals who reviewed earlier drafts of this paper. These individuals include John Centra (ETS), Jeannette Fleischner (Columbia University), Donald Powers (ETS), Lorrie Shepard (Univerity of Colorado, Boulder), Warren Willingham (ETS), John Winterbottom (ETS), and the members of the Handicapped Research Studies Group. Appreciation is also expressed to the College Board Joint Staff Research & Development Committee and the Graduate Record Examinations Board Research Committee.

Table of Contents

Handicapped Students in General 4

Level of Performance 4

Reliability 7

Validity 7

Summary 8

Learning Disabled Students 9

Level of Performance 9

Reliability 12

Validity 13

Summary 15

Physically Handicapped Students 16

Level of Performance 16

Reliability 17

Validity 17

Summary 17

Hearing Impaired Students 18

Level of Performance 18

Reliability 20

Validity 21

Summary 22

Visually Impaired Students 23

Level of Performance 23

Reliability 26

Validity 27

Summary 28

Conclusion 29

References 34

The College Board, Educational Testing Service (ETS), and the Graduate Record Examinations (GRE) Board are collaboratively sponsoring a four-year research program on admissions testing and handicapped people. The program is meant to (a) respond to issues raised by Section 504 of the Rehabilitation Act of 1973 and by the National Research Council's Report of the Panel on Testing of Handicapped People (Sherman & Robinson, 1982), and (b) ensure that the testing experience and scores on tests administered by ETS to handicapped individuals are as comparable as possible to those of nonhandicapped examinees. In line with this dual purpose, three research and development objectives were set. These were to (a) develop an improved base of information and test data concerning the characteristics of handicapped populations, (b) evaluate and improve, wherever possible, the accuracy of assessment for handicapped people, and (c) evaluate and enhance, wherever possible, the fairness and comparability of tests for handicapped and nonhandicapped examinees (see Bennett & Ragosta, 1984, for a more complete description of the research program).

To help satisfy the research program's first objective--developing an improved base of information and test data concerning the characteristics of handicapped students--two background papers were produced. The first paper, A Research Context for Studying Admissions Tests and Handicapped Populations (Bennett & Ragosta, 1984), focused upon the characteristics of the four largest disability groups taking admissions tests: students with learning disabilities, physical handicaps, hearing impairments, and visual impairments. The second, and present,

paper selectively documents existing research on the performance of these four groups with respect to admissions tests and other similar measures. In addition, the psychometric characteristics of these measures when used with handicapped examinees are reported.

Studies cited in the paper were selected on the basis of several criteria. One criterion was the similarity of the measures to college and graduate admissions tests. Cognitive ability tests and academic achievement measures were included because their content and format is similar to admissions tests. A second criterion was the similarity of the examinees to college and graduate applicants. Studies of high school and college-age students were particularly sought, though research on younger subjects was sometimes cited when little other relevant work could be located. A final criterion was the quality of research. Efforts were made to consider only those studies with samples large enough to offer dependable results (i.e., more than 40 subjects). However, because samples used in research on handicapped pupils appear to be typically small, this criterion was occasionally relaxed when little dependable data could be found.

In addition to sample size, efforts were made to include only studies for which comparable data for nonhandicapped subjects were available. Comparable data for nonhandicapped subjects are necessary to render research results interpretable. For example, the poor SAT performance of deaf students attending an institution has limited meaning without knowledge of the

performance of the institution's general student population; deaf student performance may be poor but so may the performance of the school's students generally, suggesting the possibility that poor performance is typical of those attending the institution and not necessarily of the deaf population as a whole.

The requirement for comparability was satisfied by individual studies that compared data for a sample of handicapped subjects from a particular institution or geographical area with data for an appropriate control group from the same institution or area; by individual studies that compared data for a national sample of disabled subjects with normative data for a test or other data for a nationwide cross-section of the population; or by comparisons of data from a series of similar studies of handicapped subjects with data from a series of studies of nondisabled people, each investigation being based on a sample from a particular institution or geographical area. It is recognized that this requirement does not guarantee comparable data (e.g., handicapped and nonhandicapped groups within the same institution may be selected using different decision rules), but merely enhances the likelihood of obtaining groups that are similar on important characteristics. Also, as with the sample size criterion, the comparability requirement could not always be applied without excluding all studies covering a given topic and was, therefore, occasionally relaxed.

The research results presented in the paper are discussed first in terms of handicapped students as a general group and then in terms of each of the four disability groups most commonly taking admissions tests. This approach was chosen to permit a

general picture to emerge as well as to allow for the examination of differences among the specific disability groups. For both general and specific groups, information is presented on level of test performance, reliability, and validity.

Handicapped Students in General

Level of Performance

Four studies present data relevant to the test performance of handicapped students in general. Two of these studies concern the Scholastic Aptitude Test (SAT), one the American College Testing Program (ACT) Assessment, and one the Graduate Record Examinations (GRE).

Ragosta and Nemceff (1982) reported the scores of all candidates taking special administrations of the SAT in 1981. Special administrations included braille, large-type, cassette, extra time, the use of a reader or recorder, or some combination of these accommodations (including regular-type with extra time). The mean scores for the approximately 6,800 pupils taking special administrations were 383 on the Verbal section (SAT-V) and 429 on the Mathematical section (SAT-M) for males, and 373 (SAT-V) and 395 (SAT-M) for females, respectively. During 1981, nonhandicapped males scored 430 (SAT-V) and 492 (SAT-M) while females scored 411 and 443, respectively (Ragosta & Nemceff, 1982). With the exception of the performance of females on the SAT-V, the scores of handicapped students were appreciably below (by close to one-half standard deviation) those of the general SAT-taking population.

An investigation of the performance of handicapped students

who elected to take both timed and untimed versions of the standard SAT was conducted by Centra (1983). Centra's sample of over 1,400 pupils constituted a subset of the students taking special administrations included in the Ragosta & Nemceff (1982) study described above. Centra found the mean timed score to be 344 and the mean untimed score to be 401 for the Verbal section of the test. For the Mathematical section, scores were 400 for the timed administration and 459 for the untimed administration. Scores from the timed administration were appreciably different (over .5 standard deviations) from those of all high school seniors taking the SAT during the same period (424 to 427 for SAT-V, 466 to 467 for SAT-M, Educational Testing Service, 1982). In contrast with Ragosta and Nemceff's (1982) results, scores from the untimed administration were only marginally different (less than .25 standard deviations) from the performance of high school seniors.

Centra found handicapped students to post appreciable score gains on the modified test over the timed administration. Gains averaged between 30 and 38 points after typical practice, growth, and measurement error effects had been taken into account. According to Centra, these gains were greater than those reported for nonhandicapped students given extra time (Hale, 1982), supporting the notion that time limits put disabled students at a disadvantage relative to the general test-taking population.

Aside from the College Board's Scholastic Aptitude Test, data have been reported by the American College Testing Program for the ACT Assessment. Maxey and Levitz (1980) reported performance data for a sample of approximately 10,000 disabled

students taking the standard version of the ACT Assessment on a national test date during the years 1975-76, 1976-77, and 1978-79. According to these investigators, disabled students consistently achieved a composite score about a half-standard deviation lower than the national test norms (mean scores for handicapped students were 15.4, 16.1, and 16.0; scores for others were 18.3, 18.4, and 18.6). Data for the years 1979 and 1980 show similar differences (Maxey, 1982). These performance differences are consistent with ACT data suggesting that handicapped students earn lower first year college grades (Maxey & Levitz, 1980).

The final study relevant to handicapped students in general was reported for the standard version of the Graduate Record Examinations Aptitude Test (GRE) by Wild (1982). For the 1981-1982 testing year, the mean scores of over 2,000 disabled examinees were 487 for the GRE-Verbal, 510 for the GRE-Quantitative, and 513 for the GRE-Analytical. Scores for all examinees taking the GRE during the same period were 501, 539, and 539, respectively. The differences between the performance of handicapped and nondisabled students on the GRE (less than .25 standard deviations) were not as discrepant as the differences generally evidenced between those groups on undergraduate admissions tests. The greater similarity in scores on the GRE may be due to a more select population of handicapped students applying to graduate school.

Reliability

Little data on the reliability of admissions tests for handicapped students in general is available (Bennett, 1983). A preliminary idea of the reliability of admissions tests is given by the Centra (1983) study described above. Centra reported correlations between the SAT scores that disabled candidates obtained in timed and untimed administrations. Correlations of .79 and .85 for the SAT-V and SAT-M, respectively, were obtained using a sample of over 1,800 subjects. These correlations necessarily underestimate the parallel-form reliability of the SAT for handicapped examinees as the two test administrations differ substantially in speededness as well as in other variables (Educational Testing Service, 1980). Still, only the SAT-V correlation was appreciably lower than the corresponding correlation between SAT scores in standard administrations for national samples of candidates: .88 to .91 for SAT-V and .85 to .88 for SAT-M (Donlon & Angoff, 1971).

Validity

As for reliability, little data on the validity of admissions tests for handicapped students in general appear to exist (Bennett, 1983). Data for the ACT Assessment from the 1976-1977 Basic and Standard Research Service files showed approximately equivalent correlations between the ACT Composite and college grade point average (GPA) in both handicapped (.46) and nonhandicapped (.44) student populations (Maxey & Levitz, 1980). For both groups, the combined ACT Composite score and self-reported high school average predicted college grades equally well (multiple R = .59). When actual and predicted

grades were compared, a slight underprediction occured for handicapped students (actual grade median = 2.37, predicted = 2.35) while a slight overprediction occured for the general population (actual grade mean = 2.54, predicted = 2.58). These prediction errors were so small as to be of little practical significance.

Summary

Data from both the SAT and ACT Assessment indicate that handicapped students as a group perform about one-half standard deviation below national norms. Lower performance is most evident for those handicapped students taking standard administrations and those taking only modified tests. Examinees taking both timed and untimed tests show scores only marginally different from national norms on the untimed test. Disabled students taking the standard version of the GRE also show only marginal differences in performance from national test norms (less than .3 standard deviations).

Data on the reliability and validity of admissions tests for handicapped students as a group are largely absent. Preliminary data for the SAT show no dependable differences in precision between handicapped and nondisabled populations. Data from the American College Testing Program support the validity of the ACT Assessment as an equivalent predictor of college grade point average for both disabled and nonhandicapped students.

Learning Disabled Students

Level of Performance

The most relevant data on the level of performance of learning disabled examinees on ability tests come from studies of the Scholastic Aptitude Test (SAT). In an investigation of over 1,100 candidates with learning disabilities who chose to take the SAT in both a timed as well as an untimed administration between 1979 and 1982, the candidates' mean SAT-V score was 338 and their mean SAT-M score was 393 in the standard administration; the corresponding means in the untimed administration were 394 and 452, respectively (Centra, 1983). The means for the standard administration were substantially below (by over .5 standard deviations) those for all high school seniors who took the SAT during the same period (424 to 427 for SAT-V, 466 to 467 for SAT-M, Educational Testing Service, 1982), whereas the means for the untimed administration were closer to those for high school seniors (less than .3 standard deviations away).

In contrast to Centra's study, which was limited to candidates who took the SAT in both standard and special administrations, markedly different results were obtained in an investigation that included all 4,044 candidates with learning disabilities who took the SAT in special administrations between 1979 and 1981 (Ragosta & Nemceff, 1982). (About a quarter of the examinees in the Ragosta & Nemceff study were also in the Centra investigation.) The mean SAT-V score was 372 for males and 353 for females, and the mean SAT-M score was 417 for males and 377 for females. These means were well below (at least .5 standard

deviations) those for nonhandicapped students who took the SAT during 1981 (430 on SAT-V and 492 on SAT-M for males, and 411 and 443, respectively, for females, Ragosta & Nemceff, 1982). The difference between the two investigations is presumably due to the atypical nature of the learning disabled candidate subgroup that elected to take the SAT in a standard as well as an untimed administration.

In a study of freshmen in a small college, the mean SAT-V score was 352 and the mean SAT-M score was 382 for 55 learning disabled students, whereas the corresponding means for the 111 nonhandicapped students were 411 and 405 (Jones & Ragosta, 1982). The mean difference for SAT-V was statistically significant and appreciable (over .5 standard deviation units), but the mean difference for SAT-M was not significant (less than .3 standard deviation units). It is unknown what proportion of the SAT scores for the learning disabled students came from special administrations.

Data on the performance of learning disabled students have also been reported for individually-administered tests of cognitive ability such as the Wechsler Adult Intelligence Scale (WAIS) (Wechsler, 1955) and Wechsler Intelligence Scale for Children (WISC) (Wechsler, 1949). In an investigation of students enrolled in a large university (Cordoni, O'Donnell, Raminiah, Kurtz, & Rosenshein, 1981), learning disabled and nonhandicapped students differed appreciably (by about .5 standard deviations) on their mean Full Scale WAIS IQ: the means were 108 for the former and 115 for the latter. (Verbal and

Performance IQ data were not reported.) The two groups also differed significantly and substantially in their means on 5 of the 11 WAIS subtests--Information, Similarities, Vocabulary, Digit Span, and Digit Symbol, with the means for the learning disabled students being consistently lower.

In a study of 14 year old males (Ackerman, Dykman, & Peters, 1977), the mean Full Scale IQ and mean Verbal IQ on the WISC were significantly and appreciably lower (by at least .6 standard deviations) for the learning disabled than for the nonhandicapped pupils: 103 vs. 112 for Full Scale IQ and 101 vs. 112 for Verbal IQ. The mean Performance IQ was not significantly different for the two groups.

Aside from overall performance, the subtest and subscale performance of learning disabled students has been investigated. A review of a large number of WISC investigations found that learning disabled children (specifically, those with reading disabilities) typically had their highest mean scores on the Block Design, Object Assembly, and Picture Completion subtests and their lowest mean scores on the Digit Span and Coding subtests; the subtest means for nonhandicapped children followed no consistent pattern (Rugel, 1974). In the Ackerman et al. study, the means on 4 of the 11 WISC subtests--Arithmetic, Coding, Information, and Digit Span--were also significantly and appreciably lower for the learning disabled males. (Two of these subtests, Information and Digit Span, and the WAIS counterpart [Digit Symbol] of a third subtest, Coding, also differentiated learning disabled and other college students in the previously described study by Cordoni et al.) These same four WISC subtests

comprise the "ACID" cluster (an acronym composed of the first letter of each subtest name) reported to be clinically associated with learning disability in children (Sattler, 1982).

With respect to subscale performance, an extensive review of WISC-R research found that both the Verbal IQ-Performance IQ discrepancy and subtest scatter were slightly larger for children with learning disabilities than for nonhandicapped children (Kaufman, 1981). However, differences in the amount of scatter and subscale discrepancy were too small to be of diagnostic value.

Pertinent data on the level of performance of learning disabled examinees on achievement tests are sparse. In the Ackerman et al. (1977) study of 14 year old males already cited, the means for all three scales on the Wide Range Achievement Test (WRAT) (Jastak & Jastak, 1965)--Reading, Spelling, and Arithmetic --were appreciably lower for the learning disabled than for the other students.

Reliability

Useful data on the reliability of cognitive tests for learning disabled examinees are limited. The best data come from the Centra (1983) study described earlier. This investigation reported the correlations between the SAT scores that 1,400 learning disabled candidates obtained in standard and special administrations: .76 for SAT-V and .84 for SAT-M. Because the two test administrations materially differ in time limits and other variables (Educational Testing Service, 1980), these correlations necessarily underestimate parallel-form reliability.

Nonetheless, only the SAT-V correlation was appreciably lower than the corresponding correlation between SAT scores in standard administrations for national samples of candidates: .88 to .91 for SAT-V and .85 to .88 for SAT-M (Donlon & Angoff, 1971).

McCullough and Zaremba (1979) reported findings about the internal-consistency reliability of the Woodcock Reading Mastery Test (Woodcock, 1973) and the KeyMath Diagnostic Arithmetic Test (Connolly, Nachtman, & Pritchett, 1971) for a sample of 12 to 17 year old males. The Coefficient Alpha reliabilities of the total scores on these tests were similar for the learning disabled and the other pupils in the sample: .88 for the learning disabled and .92 for the others on the Woodcock Reading Mastery Test, and .94 and .97, respectively, on the KeyMath Diagnostic Arithmetic Test.

Validity

The most pertinent findings about the validity of cognitive tests for learning disabled examinees stem from the Jones and Ragosta (1982) study of SAT scores for freshmen attending a small Massachusetts college. SAT-V and SAT-M had similar correlations with first-semester grade-point average for 55 learning disabled and 111 other students: for the former group, .32 for SAT-V and .29 for SAT-M; for the latter group, .34 and .28. However, a regression equation based on the SAT data for the nonhandicapped students systematically overpredicted the grades for the learning disabled students (i.e., promised higher grades than actually achieved). The amount of overprediction, though statistically significant, was of little practical import (.27 of a grade point).

Factor analyses of the WISC-R (Wechsler, 1974) for children

of elementary-school age also provide some insight into the validity of cognitive tests for learning disabled pupils. Several investigations of learning disabled examinees (Blaha & Vance, 1979; Naglieri, 1981; Petersen & Hart, 1979; Schooler, Beebe, & Koepke, 1978; Wallbrown & Blaha, 1979) generally found the same factor structure that had been indentified in the WISC-R standardization sample (Gutkin & Reynolds, 1981; Kaufman, 1975; Raminiah, O'Donnell, & Ribich, 1976; Reynolds & Gutkin, 1980; Silverstein, 1977; Wallbrown, Blaha, Wallbrown, & Engin, 1975). This structure consisted of a large verbal comprehension factor (mainly defined by the Information, Similarities, Vocabulary, and Comprehension subtests); a large perceptual organization factor (primarily defined by the Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Mazes subtests); and oftentimes a smaller, less stable, freedom from distractibility factor (usually defined by the Arithmetic, Digit Span, and Coding subtests).

Factor analyses of the Woodcock Reading Mastery Test (Woodcock, 1973) and the KeyMath Diagnostic Arithmetic Test (Connolly et al., 1971) were conducted by McCullough and Zaremba (1979). In their study of 12 to 17 year old males described above, the factor structure of the Woodcock Reading Mastery Test was the same for the learning disabled and the other students: one factor (defined by all the subtests) emerged. However, the factor structure of the KeyMath Diagnostic Arithmetic Test differed for the two groups. Two factors were identified for learning disabled pupils, an operations factor (defined by the

Addition, Subtraction, Multiplication, and Division subtests) and a content and application factor (defined by most of the other subtests: Numeration, Fractions, Geometry and Symbols, Word Problems, Missing Elements, Money, Measurement, and Time). Only one factor (defined by all the subtests) emerged for the other subjects.

Summary

With the exception of those who take both standard and untimed administrations, learning disabled examinees generally perform well below (by at least .5 standard deviations) nonhandicapped examinees on the Scholastic Aptitude Test. Similar performance differences (at least .5 standard deviations) are found for this group on individually-administered tests, such as the Wechsler Intelligence Scales (Full Scale and Verbal Scale), and the Wide Range Achievement Test. Studies of intra-test performance on the Wechsler Scales suggest that learning disabled students evidence only slightly larger Verbal-Performance discrepancies and subtest scatter, but consistently poorer performance on the "ACID" cluster, than nondisabled students.

Little data exist on the reliability of cognitive tests for learning disabled students. Prelimary data on admissions tests do not show any consistent differences in measurement precision between the learning disabled and general student populations. Achievement tests, such as the Woodcock Reading Mastery Test and the KeyMath Diagnostic Arithmetic Test, have also shown equivalent precision across populations.

The validity of cognitive tests for learning disabled

students has been most thoroughly investigated with respect to the WISC-R. Studies have repeatedly found the factor structure of this test when used with learning disabled examinees to mirror that found in the standardization sample. Results from a single study of the SAT suggest similar levels of predictive power for the test across populations. Finally, a single investigation into the structure of achievement tests produced equivalent factor structures for the Woodcock but not for the KeyMath.

Physically Handicapped Students

Level of Performance

The best data on the level of performance of physically handicapped examinees on admissions tests come from the Centra (1983) and the Ragosta & Nemceff (1982) studies of the SAT. For 96 candidates with physical disabilities taking the SAT in both timed and untimed administrations, the SAT-V mean was 398 and the SAT-M mean was 428 in the standard administration; the corresponding means were 446 and 486 in the untimed administration (Centra, 1983). Means from the timed administration were somewhat below (i.e., within .35 standard deviation units) while those from the untimed test were above (i.e., up to .19 standard deviations) those for all high school seniors (424 to 427 for SAT-V, 466 to 467 for SAT-M) (Educational Testing Service, 1982).

Similarly, for all 568 candidates with physical disabilities taking the SAT in special administrations (approximately a fifth of these examinees were in the Centra investigation), the SAT-V means were 424 for males and 410 for females, and the SAT-M means were 460 for males and 410 for females (Ragosta & Nemceff, 1982).

These sets of means are close (within .3 standard deviations) to the means for nonhandicapped students (430 on SAT-V and 492 on SAT-M for males, and 411 and 443, respectively, for females, Ragosta & Nemceff, 1982).

Reliability

Useful data about the reliability of ability tests for physically handicapped examinees are extremely limited. The most relevant data come from the Centra (1983) study, which reported the correlations between the SAT scores for 129 physically handicapped examinees in standard and untimed administrations. The correlations were .83 for SAT-V and .86 for SAT-M. These lower-bound estimates of parallel-form reliability were similar in level to the retest reliability of the SAT in standard administrations for national samples of examinees (.88 to .91 for SAT-V and .85 to .88 for SAT-M, Donlon & Angoff, 1971).

Validity

No relevant data could be located on the validity of cognitive tests for physically handicapped examinees.

Summary

Data on both the test performance of physically handicapped people and the psychometric characteristics of tests used with them are extremely limited. The little data that do exist suggest that these individuals perform relatively close to national norms on standard and on untimed administrations of the Scholastic Aptitude Test (i.e., within .35 standard deviation units). Preliminary reliability data suggest that the SAT measures with equal precision for physically handicapped and nondisabled groups. No validity data could be located.

Hearing Impaired Students

Level of Performance

Ragosta and Nemceff (1982) studied over 250 hearing-impaired candidates taking the SAT in nonstandard administrations between 1979 and 1981. The mean SAT-V scores were 301 for males and 299 for females, and the mean SAT-M scores were 403 for males and 381 for females. The means for hearing-impaired students were well below the means (between .5 and 1.2 standard deviations) for nondisabled students (430 on SAT-V and 492 on SAT-M for males, and 411 and 443, respectively, for females, Ragosta & Nemceff, 1982).

Jones and Ragosta (1982) studied 108 hearing-impaired students attending a state university in California. These investigators found means for the hearing impaired sample to be 355 for SAT-V and 413 for SAT-M. Corresponding means for the 259 hearing students attending the same school were 455 and 480, respectively. Differences in means between the two groups were statistically significant and appreciable (between .6 and 1.1 standard deviation units). In addition, Jones and Ragosta found mean SAT-V scores to increase with delay in onset of deafness (319 for those with onset at birth, 383 with onset in preschool, and 466 with onset at age 5 and over).

Additional information on the cognitive test performance of deaf and hearing-impaired students comes from studies using nonverbal, individually-administered tests such as the Performance Scales of the WISC (Wechsler, 1949), WISC-R (Wechsler, 1974), and WAIS (Wechsler, 1955). Schildroth (1976)

reported the nonverbal IQ scores of more than 21,000 students enrolled in programs for the hearing impaired in the 1970 to 1971 school year. The mean nonverbal IQ for all hearing impaired students was 95.6. For the subsample of hearing impaired children with no additional handicapping condition, the mean nonverbal IQ was 100.1. Students with at least one additional handicapping condition registered a mean score of 86.5. Schildroth's data are supported by Meadow (1980) who, in reviewing several studies, concluded that deaf children usually score within the average range on the Wechsler Performance Scales.

Scores on both the WAIS Performance and Verbal Scales were reported by Ross (1970) for deaf students age 16-21. Testing was accomplished using simultaneous manual and verbal communication with responses accepted in either mode. Ross found that the mean Performance IQ was 106. The mean verbal IQ was 72, a score well below (by almost two standard deviations) the performance of the general population.

Studies of the subtest performance of deaf children suggest that these pupils get their best scores on the Object Assembly and Block Design subtests of the WISC-R Performance Scale (Vonderhaar & Chambers, 1975; Hess, 1960). They seem to perform significantly more poorly than their hearing peers on the Coding and Picture Arrangement subtests (Anderson & Sisco, 1977).

Data on the achievement test performance of deaf students come from the 1974 national standardization of the Special Edition of the 1973 Stanford Achievement Test for Hearing Impaired Students (Trybus & Karchmer, 1977). Trybus and Karchmer

reported the median reading comprehension score at the test's highest age level (20+ years) to be at grade level 4.5, with only 10 percent of hearing-impaired 18 year-olds scoring at an 8th-grade level. In mathematics computation, the median score at age 20+ was just below 8th grade level, with only the top ten percent of hearing-impaired students scoring at the level of hearing students. According to Trybus and Karchmer, these results are consistent with data obtained in 1969 and 1971 from samples of 12,000 and 19,000 children, respectively.

Reliability

Reliability data on the SAT for deaf candidates is reported by Centra (1983). For 34 candidates with hearing impairments who took the SAT in both standard and untimed administrations, estimates of reliability were .89 for the SAT-V and .79 for the SAT-M. Because of differences in test administration, these correlations underestimate parallel-form reliability. Compared to a national sample of nonhandicapped candidates (.88 to .91 for SAT-V; .85 to .88 for SAT-M) (Donlon & Angoff, 1971), the reliability of SAT-V for deaf examinees is equivalent to that for the general population while the reliability of the SAT-M is only slightly lower. Given the small number of deaf students used to generate these reliability estimates, the above comparisons should be viewed with caution.

In addition to the Scholastic Aptitude Test, reliability data are reported for the Stanford Achievement Test for Hearing Impaired Students (SAT-HI) by Jensema (1978). Internal consistency coefficients for the SAT-HI subtests ranged from a

low of .64 to a high of .96 with most estimates in the .80s. These estimates are roughly comparable to reliabilities typically found for the Stanford Achievement Test when used with nondisabled students. Subtest reliabilities for Intermediate Form E range from .42 to .84, with most values falling in the .60s and .70s for fifth grade students (Gardner, Rudman, Karlsen, & Merwin, 1982).

Validity

The most pertinent findings about test validity for deaf pupils come from Jones and Ragosta's (1982) study of the Scholastic Aptitude Test (SAT) in a California state university. These researchers found correlations with freshman grade-point average of .14 and .41 for SAT-V and SAT-M, respectively, for 60 deaf students, and .38 and .32 for SAT-V and SAT-M for 140 hearing examinees. The SAT validity coefficients for the two groups were not significantly different, though the difference between the SAT-V coefficients approached statistical significance (p=.059) and warrants further investigation. When combined with high school grade point average, the SAT predicted freshman performance in the hearing impaired and hearing populations equivalently.

Research on the validity of the Gallaudet Entrance Exam is reported by Greenberg and Greenberg (1971). The exam, used at Gallaudet College, consisted of 13 tests including some commercial measures and some developed especially for deaf students. The tests were selected or designed to measure subject matter knowledge (e.g., algebra, vocabulary), special skills (e.g., writing, paragraph arrangement), and academic potential

(e.g., concept formation). Of the 13 tests, the best predictors of college performance in the Gallaudet class of 1965 were verbal measures: vocabulary and writing. Each measure correlated .44 with college grade point average. Similar correlations were observed for the following year's class (Greenberg & Greenberg, 1971).

Summary

The available research on the performance of hearing impaired students suggests that these pupils perform well below the general population on the Scholastic Aptitude Test, on the Verbal portion of the Wechsler Adult Intelligence Scale, and on reading comprehension and mathematics computation sections of the Stanford Achievement Test for Hearing Impaired Students. The discrepancy between the performance of the two populations is consistently more pronounced on the verbal portions of these measures than on the math or performance sections, a finding consonant with the well-documented natural language deficiencies of deaf individuals (Meadow, 1980).

Data on the reliability of admissions tests for deaf students are largely lacking. Data from one study show no dependable differences in measurement precision between hearing impaired and general student groups. Data for the Stanford Achievement Test also show roughly comparable reliability values across populations.

As is the case for reliability, validity data are largely absent. A preliminary study conducted at a single institution found the SAT combined with high school grades to predict college

performance equivalently in hearing-impaired and hearing populations. In addition, studies conducted at Gallaudet College showed the best predictors of success at that institution to be verbal measures.

Visually Impaired Students

Level of Performance

Data on the admissions test performance of visually-impaired people come from studies of the Scholastic Aptitude Test and ACT Assessment. Ragosta and Nemceff (1982) reported scores for almost 1,500 visually impaired candidates taking nonstandard SAT administrations between 1979 and 1981. The mean SAT-V scores were 420 for males and 411 for females and the mean SAT-M scores were 459 and 421 for males and females, respectively. These scores were only slightly different (within .3 standard deviations) from those for nondisabled students taking the test during 1981 (430 on SAT-V and 492 on SAT-M for males, and 411 and 443, respectively, for females, Ragosta & Nemceff, 1982).

In his investigation of handicapped candidates taking the SAT in both timed and untimed administrations, Centra (1983) found the mean SAT-V and SAT-M scores for some 177 visually impaired candidates to be 371 and 432, respectively, in standard administrations. Corresponding means for the nonstandard, untimed administrations were 433 for the SAT-V and 494 for the SAT-M. Only the score from the SAT-V standard administration was appreciably below (by .5 standard deviation units) that achieved by high school seniors during the same period (424 to 427 for SAT-V, 466 to 467 for SAT-M, Educational Testing Service, 1982).

The American College Testing Program (ACT) reported on the performance of 412 visually impaired candidates taking special test administrations in the 1979-1980 academic year (Maxey & Levitz, 1980). On English, Mathematics, Social Studies, and Natural Sciences tests, visually handicapped students achieved mean scores of 16.1, 16.6, 17.7, and 19.7, respectively. Nondisabled students scored comparably (i.e., within .1 to .4 standard deviation units), achieving means of 17.9, 17.4, 17.2, and 21.0, respectively, on the same measures. The mean composite score for the visually impaired was 17.7, comparing favorably with a composite of 18.5 for the nonhandicapped (i.e., a difference of less than .2 standard deviations).

In addition to performance on admissions tests, the performance of visually impaired people on the Verbal Scales of the WISC (Wechsler, 1949) and WAIS (Wechsler, 1955) has been investigated. In general, the scores of visually impaired people appear equivalent to those of the sighted population (Vander Kolk, 1977; Jordan & Felty, 1968; Smits & Mommers, 1976; Tillman, 1967). Similar results have been found using the Hayes-Binet and Interim Hayes-Binet (Hayes, 1941). Characteristically, these studies have also found higher than average scores on Digit Span for visually impaired examinees and lower than average performance on the Comprehension subtest. Finally, no differences between the performance of blind and partially sighted examinees appear evident (Vander Kolk, 1977).

While the performance of visually impaired people appears equivalent to the general population on academic ability tests,

data on achievement test performance seems more discrepant. For example, Ozias (1975) reported that three separate investigations found blind children well below the mathematics achievement level of their sighted peers. Hallahan and Kauffman (1978), in a review of the characteristics of visually impaired children, concluded that both partially sighted and blind children achieve more poorly than their sighted peers when matched on mental age. Finally, the Special Assessment of Visually Impaired Students of the Florida Statewide Assessment Program (Florida Department of Education 1981), reported that the achievement of visually disabled students tested in 1979 and 1980 was well below the achievement of regular students.

The cause of the reported discrepancies between the ability and achievement test performance of visually impaired students is not immediately clear. Such discrepancies may be due to differences in the segments of the visually impaired population that take ability and achievement tests (e.g., a more select segment may take ability tests), the use of achievement tests with content or administrative procedures that are less appropriate for assessing the skills of visually impaired students (e.g., group versus individually administered measures), or a real difference between general cognitive ability and accumulated academic knowledge.

Limited evidence that modifications in achievement test content and procedure affect the performance of visually impaired students was provided by Trismen (1967). In a study designed to produce an equivalent form of the Sequential Tests of Educational Progress for blind students, this investigator found the

performance of over 200 blind students to be roughly comparable to national norms when time limits had been relaxed and unsuitable items removed.

Reliability

Reliability data on the Scholastic Aptitude Test for visually impaired candidates is reported by Centra (1983). For 218 candidates who took the SAT under both timed and untimed conditions, estimates of reliability were .83 for the SAT-V and .84 for the SAT-M. Because of the differences in test administration, these correlations underestimate parallel-form reliability. They are, nevertheless, comparable to the reliabilities reported for a national sample of sighted candidates: .88 to .91 for SAT-V and .85 to .88 for SAT-M (Donlon & Angoff, 1971).

Tillman (1973) reported the reliability results from several studies of the WISC Verbal Scale when used with blind students. Internal consistency reliabilities on the WISC subtests ranged from .71 to .83. For the Verbal Scale, the split-half reliability was estimated to be .89 and the test-retest reliability was .91. These results are generally comparable to those reported for the WISC standardization sample. Split-half reliabilities for that sample range from .50 to .91 for the Verbal Scale subtests and .88 to .96 for the Verbal Scale IQ (Wechsler, 1949).

The test-retest reliability of the Interim Hayes-Binet used with visually handicapped students was reported by Hayes (1950) to be .90. Preliminary estimates of the split-half reliabilities

for the new Perkins-Binet were reported by Coveny (1972) to be .96 for braille readers on Form N and .94 for print readers on Form U. These estimates are generally comparable to those found in uses of the Stanford-Binet with nondisabled students, for which alternate-forms reliabilities range from .83 to .98 (Terman & Merrill, 1973).

Reliability coefficients for several achievement tests modified for use with blind students and standardized on students at Perkins Institution are reported by Bauman and Hayes (1951). Retests with forms D-H of the Stanford Achievement Tests yielded coefficients from .87 to .95. Form R of the Metropolitan Achievement Test produced a tetrachoric correlation of .98 with Form S taken two years later at Perkins. These reliabilities are roughly comparable to those typically found for the general population. For the nondisabled population, retest reliabilities for the Stanford Achievment Tests range from .87 to .93 on Intermediate 1 Forms E and F (Gardner, Rudman, Karlsen, & Merwin, 1982). For the Metropolitan, representative internal consistency estimates range from .79 to .93 for the Advanced 1 when used with eighth grade students (Farr, Prescott, Balow, & Hogan, 1978).

Validity

Very little data on the validity of cognitive tests for blind or partially sighted students exist. Lewis (1957) investigated the validity of the Interim Hayes-Binet using academic achievement as one validity criterion. She found correlations of .45, .46, and .53 between the Interim Hayes-Binet and grade averages in elementary, junior high, and high school, respectively. These results suggest a moderately strong

relationship between the test and school grades.

Coveny (1973) undertook a predictive validity study of the Perkins-Binet (PB) and the WISC Verbal Scale as predictors of the academic achievement of visually handicapped children. A multiple regression analysis showed both the PB and WISC to predict teacher grades and Stanford Achievement Test subscores. Percentages of variance (r^2) accounted for by PB subtests ranged from .08 to .23 for teacher grades, and .15 to .41 for Stanford subscores. The equivalent figures for the WISC ranged from .16 to .34 and .20 to .40, respectively.

Summary

From the available data, the performance of visually impaired students on special administrations of admissions tests appears to be close to that of the general candidate population. Only on the standard SAT-V do the scores of visually handicapped students seem appreciably lower than those of sighted students. On individually administered tests of cognitive ability, such as the WISC Verbal Scale, visually-impaired examinees achieve scores equivalent to those of the national norm group. Performance on achievement tests appears to be below national norms.

The existing research on the reliability of cognitive tests for visually impaired pupils suggests values comparable to that for the sighted population for a variety of tests including the Scholastic Aptitude Test, WISC Verbal Scale, Interim Hayes Binet, Stanford Achievement Tests, and Metropolitan Achievement Tests. Though a few studies suggest significant relationships between individually-administered cognitive tests and grades, data are

largely lacking on the validity of cognitive tests for visually impaired examinees.

Conclusion

This paper has documented the available data on the performance of handicapped people on admissions and similar tests. Additionally, information on the psychometric characteristics of these measures was reported. Though much more information needs to be collected, some tentative observations can be made from the existing data.

Table 1 lists performance differences between handicapped examinees and the general population in standard deviation (SD) units. As the table suggests, the admissions test performance of physically handicapped students and of visually impaired examinees is most similar to the nondisabled population. As noted earlier, the available data suggest visually impaired students only perform appreciably lower than national norms on the standard SAT-V. Their performance on individually-administered verbal ability tests is generally equivalent to national norms, though they appear to do more poorly on achievement measures.

Insert Table 1 about here

The admissions test performance of learning disabled individuals appears next among the specific disability groups in distance from the general population. With the exception of those who take both untimed and standard administrations, the admissions test performance of these examinees is appreciably

lower than that of their nondisabled peers (generally by at least .5 standard deviations). Similar performance differences are reported for individually administered tests such as the Wechsler Scales (Full Scale and Verbal IQ) and the Wide Range Achievment Test.

Hearing impaired students perform the least well of all groups on admissions tests. They also show performance well below that of national norms on the Verbal portion of the WAIS and on the reading comprehension and mathematics computation sections of the Stanford Achievement Test for Hearing Impaired Students. Their performance is consistently more different from the nondisabled population on verbal than on math or performance measures. The poor performance of this group on admissions tests and the effects such performance may have on college and graduate admissions, make further investigation of the fairness of admissions tests for this group (as well as for those with learning disabilities) critical.

The limited data on the reliability of admissions tests show no trustworthy differences between specific disability groups and the general test-taking population (see Table 2). The fact that these estimates were derived by correlating scores from dissimilar administrations (i.e., timed and untimed), however, suggests the need for further investigation before definitive conclusions can be offered. Finally, data from studies on several school ability and achievement tests suggest comparable measurement precision between general and visually impaired, deaf, and learning disabled examinees.

Insert Table 2 about here

The results of the small number of admissions test validity investigations are presented in Table 3. Data on the ACT Assessment support the validity of this measure as an equivalent predictor of college performance among disabled students in general and nonhandicapped examinees. Two studies of the SAT at individual institutions show no significant differences in validity coefficients for deaf and learning disabled students compared to their nonhandicapped peers. Finally, the factor structure of the WISC-R for learning disabled examinees has been consistently demonstrated to mirror that for the general population.

Insert Table 3 about here

The last observation to be made from this review relates to the paucity of relevant, dependable data. The investigations reviewed in the paper were located through extensive searches of the ERIC and Psychological Abstracts data bases, journals such as the Journal of Learning Disabilities and American Annals of the Deaf, and standard texts such as Sattler's (1982) Assessment of Children's Intelligence and Special Abilities. Because of the large amount of special education testing done in the nation's schools, reports were also obtained from several large school districts. Despite this search, little dependable data on the

cognitive test performance of handicapped students and the psychometric characteristics of tests used with these individuals were found.

The major implication of this lack of data is that defensible conclusions cannot yet be drawn about the fairness of admissions and similar tests for assessing the abilities of handicapped people. The lack of data leave unanswered many important questions that must be resolved before definitive statements can be made. For one, the lack of data leave undefined the exact nature of the subgroup of handicapped persons taking admissions tests. This subgroup may be more or less select than the subgroup of nondisabled students who take admissions tests. Differences in the nature of these two subgroups may result in estimates of test performance, reliability, and validity that differ, but that are not representative of the values that characterize the populations from which these subgroups come.

A second unanswered question relates to the cause of those differences in test performance that are determined to distinguish handicapped and nondisabled populations. Such differences may be the result of test bias, as when disparities in tested performance across groups are <u>not</u> reflected in similarly different levels of future academic accomplishment. Alternatively, tested differences may reflect real disparities in the abilities of handicapped and nondisabled people. Such differences would be expected to manifest themselves in poorer academic performance and, hence, would not imply the existence of bias in admissions tests.

To resolve these questions, further studies of test performance and of the reliability, predictive validity, and factor structure of the scholastic measures used with disabled students are badly needed. The research program being undertaken by the College Board, Educational Testing Service, and the Graduate Record Examinations Board includes many of these investigations. The collection of such information should provide a better basis for evaluating the fairness of tests for disabled individuals and for improving the admissions testing process.

References

Ackerman, P. T., Dykman, R. A., & Peters, J.B. Learning-disabled boys as adolescents--Cognitive factors and achievement. Journal of the American Academy of Child Psychiatry, 1977, 16, 296-313.

Anderson, R. J., & Sisco, F. V. Standardization of the WISC-R Performance Scale for deaf children. Washington, DC: Office of Demographic Studies, Gallaudet College, 1977.

Bauman, M. K., & Hayes, S. P. A Manual for the psychological evaluation of the adult blind. New York: The Psychological Corporation, 1951.

Bennett, R. Research and evaluation priorities for special education assessment. Exceptional Children, 1983, 50, 110-117.

Bennett, R., & Ragosta, M. A research context for studying admissions tests and handicapped populations. Princeton, NJ: Educational Testing Service, 1984.

Blaha, J., & Vance, H. The hierarchical factor structure of the WISC-R for learning disabled children. Learning Disability Quarterly, 1979, 2, 71-75.

Centra, J. Handicapped student performance on a timed vs. untimed SAT. Unpublished manuscript, Educational Testing Service, 1983.

Connolly, A. J., Nachtman, W., & Pritchett, E. M. KeyMath Diagnostic Arithmetic Test manual. Circle Pines, MN: American Guidance Service, 1971.

Cordoni, B. K., O'Donnell, J. P., Ramaniah, N. V., Kurtz, J., & Rosenshein, K. Wechsler adult intelligence score patterns for learning disabled young adults. Journal of Learning Disabilities, 1981, 14, 404-407.

Coveny, T. E. A new test for the visually handicapped: Preliminary analysis of the reliability and validity of the Perkins-Binet. Education of the Visually Handicapped, 1972, 4, 97-101.

Coveny, T. E. The Perkins-Binet and verbal WISC as predictors of academic achievement for visually handicapped children (Doctoral dissertation, George Peabody College). Ann Arbor, Mich.: University Microfilms, 1973, No. 32-622.

Donlon, T. F., & Angoff, W. H. The Scholastic Aptitude Test. In W. H. Angoff (Ed.), The College Board Admissions Testing Program: A technical report on research and development activities relating to the Scholastic Aptitude Test and Achievement Tests. New York: College Entrance Examination Board, 1971.

Educational Testing Service. Supervisor's Manual--Instructions for nonstandard administrators of the admissions testing program. Princeton, NJ: Author, 1980.

Educational Testing Service. National report on college-bound seniors, 1982. New York: College Entrance Examination Board, 1982.

Farr, R. C., Prescott, G. A., Balow, I. H., & Hogan,T. P. Metropolitan Achievement Test Advanced 1 Forms JI and KI: Teacher's manual for administering and interpreting. NY: Psychological Corporation, 1978.

Florida Department of Education Division of Public Schools. Florida Statewide Assessment Program: Visually impaired students special assessment. Tallahassee, FL: State of Florida, Department of Education, 1980.

Gardner, E. F., Rudman, H. C., Karlsen, B., & Merwin, J. C. Stanford Achievement Test Intermediate 1 Form E: Norms booklet for Fall. NY: The Psychological Corporation, 1982.

Greenberg, B. L., & Greenberg, S. H. The measurement of college potential in the hearing handicapped. American Annals of the Deaf, 1971, 116(3), 372-381.

Gutkin, T. B., & Reynolds, C. R. Factorial stability of the WISC-R for white and black children from the standardization sample. Journal of Educational Psychology, 1981, 71, 227-231.

Hale, G. Effects of extra time on SAT and Achievement Tests. Unpublished paper. Princeton, NJ: Educational Testing Service, 1982.

Hallahan, D., & Kauffman, J. Exceptional children: Introduction to special education. Englewood Cliffs, NJ: Prentice-Hall, 1978.

Hayes, S. P. Contributions to a psychology of blindness. New York: American Foundation for the Blind, 1941.

Hayes, S. P. Measuring the intelligence of the blind. In W. Donahue & D. Dabelstein (Eds), Psychological diagnosis and counseling of the adult blind. New York: American Foundation for the Blind, 1950.

Hess, W. Personality adjustment in deaf children. Unpublished doctoral dissertation, University of Rochester, 1960.

Jastak, J. F.. & Jastak, S. R. The Wide Range Achievement Test manual of instructions. Wilmington, DE: Guidance Associates of Delaware, 1965.

Jensema, C. A comment on measurement error in achievement tests for the hearing impaired. American Annals of the Deaf, 1978, 123(6), 496-499.

Jones, D. H., & Ragosta, M. Predictive validity of the SAT on two handicapped groups: The deaf and the learning disabled. (RR 82-9). Princeton, NJ: Educational Testing Service, 1982.

Jordan, J. E., & Felty, J. Factors associated with intellectual variation among visually impaired children. American Foundation for the Blind Research Bulletin, 1968, 15, 61-70.

Kaufman, A. S. Factor analysis of the WISC-R at 11 age levels between 6 1/2 and 16 1/2 years. Journal of Consulting and Clinical Psychology, 1975, 43, 135-147.

Kaufman, A. S. The WISC-R and learning disabilities assessment: State of the art. Journal of Learning Disabilities, 1981, 14, 520-526.

Lewis, L. L. Relation of measured mental ability to school marks and academic survival in the Texas School for the Blind. The International Journal for the Education of the Blind, 1957, 6, 56-60.

Maxey, J. Personal communciation, March 15, 1982.

Maxey, J., & Levitz, R. S. ACT services for the handicapped. Paper presented at the meetings of the American Association of Collegiate Registrars and Admissions Officers, New Orleans, April 1980.

McCullough, B. C,. & Zaremba, B. A. Standardized achievement tests used with learning disabled and non-learning disabled adolescent boys. Learning Disability Quarterly, 1979, 2, 65-70.

Meadow, K. P. Deafness and child development. Berkeley: University of California Press, 1980.

Naglieri, J. A. Factor structure of the WISC-R for children identified as learning disabled. Psychological Reports, 1981, 49, 891-895.

Ozias, D. K. Achievement assessment of the visually handicapped. Education of the Visually Handicapped, 1975, 76-84.

Petersen, C. R., & Hart, D. H. Factor structure of the WISC-R for a clinic-referred population and specific subgroups. Journal of Consulting and Clinical Psychology, 1979, 47, 643-645.

Ragosta, M., & Nemceff, W. P. A research and development program on testing handicapped people. (RM 82-2). Princeton, NJ: Educational Testing Service, 1982.

Raminiah, N. V., O'Donnell, J. P., & Ribich, R. Multiple-group factor analysis of the Wechsler Intelligence Scale for Children. Journal of Clinical Psychology, 1976, 32, 829-83.

Reynolds, C. R., & Gutkin, T. B. Stability of the WISC-R factor structure across sex at two age levels. Journal of Clinical Psychology, 1980, 36, 775-777.

Ross, D. R. A technique of verbal ability assessment of deaf students. Journal of Rehabilitation of the Deaf, 1970, 3(3), 7-15.

Rugel, R. P. WISC subtest scores of disabled readers: A review with respect to Bannatyne's recategorization. Journal of Learning Disabilities, 1974, 7, 48-55.

Sattler, J. M. Assessment of children's intelligence and special abilities. (2nd ed.). Boston: Allyn and Bacon, 1982.

Schildroth, A. The relationship of nonverbal intelligence test scores to selected characteristics of hearing impaired students. Paper presented at the Third Gallaudet Symposium on Research in Deafness, Gallaudet College, Washington, D.C., January 1976.

Schooler, D. L., Beebe, M. C., & Koepke, T. Factor analysis of the WISC-R scores for children identified as learning disabled, educable mentally impaired, and emotionally impaired. Psychology in the Schools, 1978, 15, 478-485.

Sherman, S., & Robinson, N. Ability testing of handicapped people: Dilemma for government, science, and the public. Washington, D.C.: National Academy Press, 1982.

Silverstein, A. B. Alternative factor analytic solutions for the Wechsler Intelligence Scale for Children-Revised. Educational and Psychological Measurement, 1977, 37, 121-124.

Smits, B. W., & Mommers, M. J. Differences between blind and sighted children in WISC verbal subtests. The New Outlook for the Blind, 1976, 70(6), 240-246.

Terman, L. M., & Merrill, M. A. Stanford-Binet Intelligence Scale: 1973 norms edition. Boston: Houghton Mifflin Company, 1973.

Tillman, M. H. The performance of blind and sighted children on the Wechsler Intelligence Scale for Children: Study I. International Journal for the Education of the Blind, 1967, 16(3), 65-74.

Tillman, M. H. Intelligence scales for the blind: A review with implications for research. Journal of School Psychology, 1973, 11(1), 80-87.

Trismen, D. A. Equating braille forms of the Sequential Tests of Educational Progress. Exceptional Children, 1967, 419-424.

Trybus, R. J. & Karchmer, M. A. School achievement scores of hearing impaired children: National data on achievement status and growth patterns. American Annals of the Deaf, 1977, 122(4), 62-69.

Vander Kolk, C. J. Intelligence testing for visually impaired persons. Visual Impairment and Blindess, 1977, 158-163.

Vander Kolk, C. J. Assessment and planning with the visually impaired. Baltimore, MD: University Park Press, 1981.

Vonderhaar, W. F., & Chambers, J. F. An examination of deaf students' Wechsler Performance Subtest scores. American Annals of the Deaf, 1975, 120, 540-544.

Wallbrown, F. H., & Blaha, J. The hierarchical factor structure of the WISC-R for reading disabled children. Multivariate Experimental Clinical Research, 1979, 4, 73-80.

Wallbrown, F. H., Blaha, J., Wallbrown, J. D., & Engin, A. W. The hierarchical factor structure of the Wechsler Intelligence Scale for Children--Revised. Journal of Psychology, 1975, 89, 223-235.

Wechsler, D. Wechsler Intelligence Scale for Children manual. New York: Psychological Corporation, 1949.

Wechsler, D. Manual for the Wechsler Adult Intelligence Scale. New York: Psychological Corporation, 1955.

Wechsler, D. Manual for the Wechsler Intelligence Scale for Children--Revised. New York: Psychological Corporation, 1974.

Wild, C. Internal memorandum on testing of handicapped persons. Princeton, NJ: Educational Testing Service, August 5, 1982.

Woodcock, R. W. Woodcock Reading Mastery Tests manual. Circle Pines, MN: American Guidance Service, 1973.

Footnotes

1. The authors are listed in alphabetical order. Majorie Ragosta drafted the sections on the hearing impaired and the visually impaired. Those on the learning disabled and physically handicapped were drafted by Lawrence J. Stricker. Kandy Elliot Bennett developed the introduction, the section on handicapped students in general, the summaries, and the conclusion, and completed the manuscript's final draft.

Table 1

Differences Between The Admissions Test Performance of Handicapped and Nondisabled Examinees[a]

Group	Median and Range of Differences in SD Units
Physically Handicapped	-.15 (-.33 to +.19)
Visually Impaired	-.16 (-.50 to +.24)
Learning Disabled	-.57 (-.80 to -.12)
Hearing Impaired	-.88 (-1.17 to -.57)

[a] Data are from studies of the SAT, GRE, and ACT Assessment described in this paper. Performance differences are expressed in standard deviation units of the nondisabled group and include results of both standard and nonstandard administrations.

Table 2

Reliability Estimates Associated with the Admissions Test Performance of Handicapped and Nondisabled Examinees[a]

Group	Reliability Estimate SAT-V	SAT-M
Nondisabled	.88-.91	.85-.88
Visually Impaired	.83	.84
Physically Handicapped	.83	.86
Learning Disabled	.76	.84
Hearing Impaired	.89	.79
Handicapped Students in General	.79	.85

[a] Data on handicapped student groups are from Centra (1983). Data on nondisabled students are from Donlon and Angoff (1971).

Table 3

Validity Estimates Associated with the
Admissions Test Performance
of Handicapped and Nondisabled Examinees

Group	Validity Estimate SAT-V	SAT-M	ACT
Jones & Ragosta (1982)			
Nondisabled	.34	.28	--
Learning Disabled	.32	.29	--
Nondisabled	.38	.32	--
Hearing Impaired	.14	.41	--
Maxey & Levitz (1980)			
Nondisabled	--	--	.44
Handicapped Students in General	--	--	.46